
Les documents électroniques dans le système d'écriture et d'édition Arcane.

Eric-Olivier Lochard

*Université Paul-Valéry
Département de mathématiques et informatique
34199 Montpellier cedex 5
eric-olivier.lochard@univ-montp3.fr*

RÉSUMÉ. Cet article décrit le modèle intellectuel, fonctionnel et interne du document implémenté dans un système opérationnel d'écriture électronique et d'édition scientifique, Arcane. Après avoir rappelé le contexte problématique de son élaboration et les a priori sur la notion de document qui l'ont influencé, on décrit le système Arcane et le paradigme qu'il implémente, puis l'équation fondatrice du modèle : document Arcane = support + enrichissements. On montre que ce modèle est particulièrement adapté pour prendre en compte opératoirement les problèmes de descriptions sémantiques, objectives ou subjectives, et la dualité forme/fond ; on explique comment il permet d'automatiser la production de livres papier aussi bien que de livres électroniques sous la forme de sites web.

ABSTRACT.

MOTS-CLÉS : Document électronique, document virtuel, édition critique, livre électronique, publication électronique, balisage, enrichissement, web sémantique, modélisation de connaissances, production de ressources électroniques

KEYWORDS:

1. Introduction

Arcane est un système informatique développé à l'usage des étudiants et des chercheurs en sciences humaines et sociales, pour produire individuellement ou collectivement d'importantes bases de documents électroniques à forte valeur ajoutée, et les publier sous forme de livres papier et/ou de livres électroniques : éditions savantes et pédagogiques, thèses, bases documentaires personnelles de recherche. C'est aussi un instrument convivial de production générique de documents balisés.

Implémentation du paradigme instrumental « Le monde selon Arcane », ce système a été imaginé et mis au point de 1992 à 2000 à Montpellier, France, dans un laboratoire de sciences humaines et sociales (SHS) du Centre National de la Recherche Scientifique ; présenté pour la première fois, en 1994, à la Maison des sciences de l'Homme à Paris, il a été lauréat du Concours national 2001 d'aide à la création d'entreprises de technologies innovantes, organisé par le Ministère de la Recherche, en option émergence.

Il est utilisé aujourd'hui par plusieurs programmes d'édition critique de correspondances du XVIII^e siècle ; d'autres chercheurs, équipes d'auteurs et doctorants l'expérimentent en grandeur réelle pour leurs propres travaux de recherche et leurs thèses, à la place des logiciels habituels de bureautique [voir annexe].

Arcane est né à un moment où le cdrom l'emportait de loin sur le web comme support idéal des productions électroniques, dans les lieux peu fréquentés de l'interdisciplinarité effective ; nourri de la culture SGML, il s'est développé au rythme lent des évolutions paradigmatiques et des projets de longue haleine que sont les éditions critiques. Aujourd'hui, certaines des questions traitées dans son développement se retrouvent dans celles que pose l'industrialisation des technologies de l'internet : architecture de la connaissance, modélisation des contenus, hiérarchisation des publications, modèle du document électronique.

Cet article va essentiellement traiter du dernier point en présentant le document électronique à travers le prisme Arcane : nous décrirons le modèle implémenté dans notre système, le « document Arcane », nous illustrerons sa mise en œuvre effective dans le travail quotidien, enfin nous évoquerons quelques perspectives encore à explorer, comme l'échange dynamique dans un réseau de publications électroniques d'un domaine donné. Avant cela, il nous faudra exposer les circonstances du développement d'Arcane et nos conceptions *a priori* de l'objet document, lesquelles expliquent les options choisies dans la réalisation de ce projet, puis présenter brièvement le système Arcane en insistant sur ce qui fait son originalité.

Avant d'entreprendre ce programme, nous devons attirer l'attention du lecteur sur l'ambiguïté du terme « éditer » qui est utilisé en français à la fois pour « publier et distribuer » (édition commerciale), et pour « préparer à la publication, annoter » (édition intellectuelle). Cette ambiguïté n'a pas de conséquence pratique dans le cas d'une édition papier puisque ces deux activités sont inséparablement liées, et que tous les acteurs de la chaîne éditoriale ont le même horizon : un livre matériel papier ; mais

dans le cas de l'édition électronique, dont une des caractéristiques est précisément la dissociation opératoire entre la rédaction (écrire et enrichir) et la publication (mettre en forme et distribuer), elle pourrait s'avérer très gênante dans cet article : on ne perdra donc pas de vue qu'« éditer » désigne explicitement ici l'ensemble des quatre opérations « écrire, enrichir, mettre en forme, publier ».

1.1. *Contexte problématique du système Arcane*

A l'origine du projet Arcane, la volonté d'un laboratoire CNRS (SHS) de tirer parti des avancées de la recherche en informatique pour ses travaux sur l'histoire des idées et de la sociabilité au XVIII^e siècle, matérialisés, en particulier, par des éditions critiques de correspondances. Le projet devait concerner à la fois le travail local, pratique et quotidien des rédacteurs, pour transformer les éditions papier en éditions électroniques, le partage d'informations comme les index et la bibliographie pour unifier et simplifier l'édition de collections d'ouvrages, et enfin la mise en réseau des publications électroniques ainsi produites dans le cadre d'un « Atlas des correspondances à l'Age moderne ».

Il s'inscrivait donc dans une double perspective : pragmatique, avec la réalisation d'outils logiciels attractifs et suffisamment aboutis pour servir d'instruments quotidiens de travail à des rédacteurs sans réelle culture ni support informatique, en lieu et place des traitements de documents habituels ; et théorique, puisqu'il supposait un travail de réflexion sur le processus de l'édition électronique dans son ensemble, de l'écriture à la publication en passant par l'architecture des données, l'élaboration d'un modèle de livre électronique comme horizon possible de l'édition, et la mise au point d'un échange dynamique entre ouvrages électroniques.

Cette double visée explique la mise en œuvre interdisciplinaire du projet : des équipes d'éditeurs travaillaient à leur rythme sur leurs projets éditoriaux, utilisant et testant les versions successives du système mises à leur disposition ; en retour, leurs observations permettaient d'imaginer et d'installer de nouveaux dispositifs opérationnels de lecture/écriture, et d'affiner la spécification du système informatique et de ses modules.

Il faut redire combien la réalisation d'un tel projet a représenté une affaire de longue haleine, puisqu'elle a toujours été directement associée à d'importants travaux d'édition échelonnés sur plusieurs années.

1.1.1. *Un instrument d'édition savante*

Le premier objectif du système consistait donc à revisiter le processus d'édition : il s'agissait de libérer les rédacteurs des procédures éditoriales mécanisables, en particulier de la mise en forme, de mettre en œuvre une architecture hypermédia de l'information dans une base de données, et par une généricité suffisante, de valoriser l'important et coûteux travail éditorial en favorisant la multipublication sous différents formats et

supports, et à destination de publics divers : une façon de privilégier la part irréductible de l'apport humain, c'est-à-dire le travail de qualité sur le contenu.

1.1.2. *Un instrument d'écriture électronique*

Un objectif parallèle et complémentaire portait (et porte toujours) sur les mécanismes de l'écriture électronique : le système devait nous permettre de tester en grandeur réelle des solutions innovantes aux nouveaux problèmes qu'elle pose.

On peut d'abord considérer l'écriture électronique comme la production de documents virtuels, c'est-à-dire de documents générés par application d'algorithmes à des éléments de contenu de base ; cette conception se focalise sur la description et l'articulation de trois notions : éléments de base, modèle sémantique de ces éléments et algorithmes de composition ; elle est étudiée, par exemple, dans le contexte du e-learning [FalquetAl04] et dans les travaux sur les documents virtuels personnalisables ([Ranwez] est une présentation générale de la problématique des DVP).

Le contexte de développement du projet Arcane nous a conduit à adopter un point de vue un peu différent, et à qualifier d'électronique non pas tant l'acte d'écrire lui-même qui reste dans le système Arcane une activité humaine et « manuelle », que l'objet final : pratiquer l'écriture électronique, c'est écrire des documents électroniques, matériels et/ou virtuels. Cette façon de voir déplace donc la problématique de l'écriture électronique vers celle du document lui-même, thème central de cet article. Nous verrons cependant que la production de documents virtuels est une problématique très présente dans le projet Arcane, notamment dans la composition de parcours de lecture intentionnelle, dans la génération de documents de travail conduite par l'utilisateur, ou même dans la web-publication.

La conception d'un système d'écriture pose d'autres problèmes, généraux ou spécifiques à l'édition savante ; comme ils ne rentrent pas directement dans le cadre de cet article, nous signalerons seulement le problème de l'architecture de l'information, qui a eu une influence notable sur la question du document dans Arcane. Cette architecture doit d'abord prendre en charge la structure graphique de l'écriture, son aspect multi-média et la dissociation contenant-contenu ; elle doit être assez simple pour permettre aux rédacteurs d'identifier sans ambiguïté la nature de l'information à rédiger, pour indexer de manière rigoureuse les documents, et pour faciliter la navigation des lecteurs dans le monde de connaissances ainsi composé ; mais en même temps, elle doit être suffisamment générique pour s'appliquer à un large éventail de projets éditoriaux.

1.1.3. *Un modèle pour la publication électronique*

Le dernier aspect important qui a contribué à la définition du modèle Arcane concerne la publication, dans ses deux acceptions : l'objet distribué (que publie-t-on ?) et le mode de distribution (comment le publie-t-on ?).

Une des raisons de concevoir Arcane était de faciliter l'édition d'une collection composée de plusieurs volumes sur le même domaine de connaissance. Comme le contenu de ces livres était évidemment enregistré dans un seul ouvrage-base de don-

nées¹, la question de sa publication électronique s'est immédiatement posée, avec la nécessité de donner un statut intellectuel et économique à cette publication : nous l'avons naturellement appelé « livre électronique ». La nécessité d'une réflexion approfondie sur le nouveau statut de ces productions nous paraissait d'autant plus indispensable qu'on pouvait déjà entrevoir les sérieux problèmes d'évaluation scientifique que la généralisation de la mise en ligne de documents sur le réseau internet allait poser. Dans le paradigme de l'édition selon Arcane, ce livre électronique (qui ne se confond ni avec le livre numérique, ni avec l'ex module électronique e-book) est un contrat de lecture, intellectuellement matérialisé par un monde de connaissances : « le monde selon Arcane » ; ce monde, intentionnellement et scientifiquement composé par l'auteur, est donné au lecteur pour qu'il le découvre et l'explore ; les documents électroniques, vecteurs privilégiés de cette lecture, sont inscrits dans un environnement hypermédia d'informations, en particulier sémantiques, qui éclairent leurs contenus et invitent le lecteur à emprunter des parcours proposés par l'éditeur.

La problématique du mode de distribution était principalement de nature économique : diminuer le coût élevé de l'édition électronique en réduisant celui de la publication, et augmenter les retours sur investissement par la multipublication. Ces deux principes conduisaient à dissocier opératoirement les activités de rédaction et celles de publication : les rédacteurs devaient pouvoir enrichir le monde de connaissances sans préjuger de la forme que prendrait la production finale, et ceci uniquement grâce à des procédures « logiques » ; ils assignaient au système d'être capable de produire automatiquement des publications de contenus adaptés à des publics variés, sur différents supports, à partir d'un même ouvrage.

Une des réponses a été d'équiper Arcane d'un mécanisme d'enrichissement original, mis à la disposition des rédacteurs pour ajouter de l'information dans l'ouvrage. C'est ce mécanisme qui donne toute son efficacité au modèle du document électronique.

1.2. *Prolégomènes concernant le document*

Ce modèle n'est pas issu d'une réflexion théorique préalable sur le document, comme les très intéressants travaux récents de Pédaque ([Pédaque03] et [Pédaque04]), dont d'ailleurs les propositions rappellent souvent Arcane. Il s'est imposé petit à petit, au cours du processus de développement interdisciplinaire du système, largement inspiré par nos travaux antérieurs sur le langage, et par quelques positions de principe préalables.

Dès 1990, dans les traces de la technologie SGML, il était déjà clair que le modèle Arcane devait permettre de représenter un document textuel comme une suite de

1. Par la suite, le terme *ouvrage* utilisé sans contexte explicite désignera à la fois la base de données dans laquelle le système Arcane enregistre les informations, et l'ensemble des informations intellectuelles sur lesquelles travaillent les utilisateurs.

caractères « algébriquement » balisée par des marqueurs attribués. De plus, la multi-publication imposait à ce balisage d'être logique. Tout cela paraît aujourd'hui naturel.

La seconde évidence était que ce modèle devait intégrer l'aspect multimédia du document : un texte, une feuille de calcul, une image, un enregistrement sonore ou vidéo.

Plus délicate est la question de l'aspect signifiant du document, en raison surtout de la nature des matériaux documentaires qui nous intéressent, et qui n'ont rien d'administratifs. Notre point de vue fondamental, énoncé de manière un peu abrupte, est qu'un « document ne parle pas » : la signification d'un document est construite par le lecteur qui produit une représentation et une interprétation liées à sa culture et à l'environnement dans lequel ce document est présenté. Un document n'est pas identifiable à une signification mais à un potentiel, une espérance de significations. Un art de l'écriture consisterait ainsi à composer un contenu en dosant deux ambitions contradictoires : faire connaître au lecteur l'interprétation de l'auteur, et lui permettre d'élaborer ses propres interprétations.

De plus, toutes ces constructions sémantiques s'organisent autour d'un invariant que le modèle du document doit représenter.

Cette conception s'applique également à l'éditeur qui voudra tôt ou tard enrichir le document de points de vue différents, complémentaires voire contradictoires, et proposer ces différentes significations au lecteur ; on pourrait ainsi imaginer un document comme un ensemble de calques empilables sur un fond : on choisit les calques à empiler en fonction de la proposition à faire au lecteur.

Cette problématique ne se pose pas pour les outils de lecture qui s'appliquent aux documents électroniques actuellement publiés, et qui, en reprenant la métaphore, n'ont pas à prendre en charge les calques : le choix des calques à empiler a été fait au moment de la publication, implicitement ou explicitement, et ils sont encapsulés dans le document identifié à une représentation.

Le problème se complique lorsqu'il s'agit de développer un système d'écriture dans lequel il faut prévoir d'enregistrer les différents points de vue de l'éditeur, le « potentiel de signification ».

La réponse à cette problématique est le fondement du modèle Arcane : un découplage effectif du support du document (l'invariant) et des enrichissements (les points de vue). Il permet, en particulier, de distinguer le document comme abstraction (le support et tous ses enrichissements, même contradictoires ou incompatibles), et sa représentation (le support et une partie déterminée et cohérente des enrichissements). Ce *principe de dualité* exprimé par l'équation « document Arcane = support + enrichissements » constitue l'originalité véritable de notre modèle ; nous y reviendrons longuement, après avoir présenté le paradigme et le système Arcane.

2. Paradigme et système Arcane

Arcane est fondamentalement un paradigme instrumental pour l'édition d'ouvrages électroniques à forte valeur ajoutée ; il propose une définition opératoire du processus dans son ensemble, une méta-architecture hypermédia pour enregistrer l'information, une méthodologie adaptée, et un instrument informatique pour le mettre en œuvre : le système Arcane.

2.1. L'ouvrage et l'édition électronique

Nous l'avons dit plus haut, le paradigme Arcane considère l'ouvrage électronique comme un contrat entre éditeur et lecteur, matérialisé par un monde de connaissances intentionnelles construit par l'éditeur et donné à explorer au lecteur, « Le monde de ... selon Arcane » ; les informations sont matérielles (« de quoi parle-t-il »), factuelles (« qu'est ce qu'on en sait »), ou subjectives (« qu'est qu'on en dit »).

L'édition est un processus coopératif, en principe sans fin, consistant à :

- définir l'architecture du monde à éditer comme instance de la méta-architecture Arcane, et prescrire les outils de représentation ;
- construire le monde des *sujets* d'intérêt du projet éditorial ;
- enrichir ce monde par des *documents* multimédias et des *relations*, pour le décrire, le structurer, l'illustrer, l'interpréter et le représenter ;
- en extraire à des moments choisis des sous-ensembles cohérents pour produire et publier des livres électroniques sous différentes formes.

2.2. La méta-architecture

Le monde selon Arcane se compose donc de sujets, de documents, de relations et d'enrichissements qui lui donnent son architecture hypergraphique ; s'y ajoutent des objets utilitaires, des structures (compositions de champs habituelles) pour typer les sujets et définir les relateurs, des vues pour présenter les objets, et des accessoires pour faciliter le travail éditorial et la lecture : parcours, dossiers, styles, modèles, scripts, etc. Un langage fondé sur le méta-langage Arcane, les objets créés et les termes définis par l'utilisateur, permet de décrire formellement ce monde.

Les documents et les enrichissements étant exposés en détail dans la suite de l'article, nous présentons brièvement ici les sujets et les relations.

2.2.1. Sujets

Les sujets constituent les centres d'intérêt du projet éditorial : personnes, lieux, ouvrages, événements, thèmes, etc. Ils sont librement créés dans l'ouvrage, comme instances d'un type qui définit leurs champs descripteurs. Du point de vue de la sé-

mantique du monde, ils constituent une partie essentielle des substantifs, et sont donc les candidats naturels à l'indexation sous toutes ses formes, directe ou par le biais des relations. Pour les lecteurs, ils sont des entrées privilégiées pour explorer l'ouvrage et pour consulter la documentation les concernant et leurs occurrences, dans des vues spécialisées.

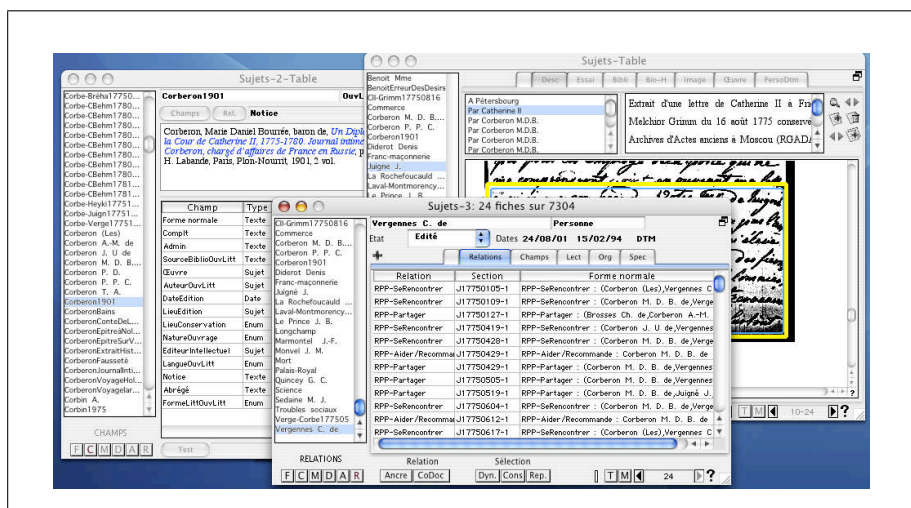


Figure 1. Trois sujets Arcane dans différentes vues.

La figure 1 présente des sujets dans trois vues différentes : la vue de ses champs descripteurs enrichis, la vue de sa documentation, ici une documentation image, et ses occurrences comme argument d'une relation.

2.2.2. Relation

Une relation est une combinaison d'objets Arcane et des valeurs typées ; elle est librement créée par instanciation d'un relateur qui détermine le nom de la relation et ses arguments, les relateurs étant préalablement définis par le responsable de l'édition, comme d'ailleurs les types de sujets.

Les relations ont trois fonctions essentielles : organiser les sujets, exprimer leurs relations « objectives » dans le contexte du projet éditorial, et formaliser le repérage d'informations pertinentes dans le matériel documentaire². Par exemple, dans les éditions de corpus de correspondances utilisées pour étudier la sociabilité au XVIII^e siècle, les relateurs vont décrire des faits, des événements, des comportements, éventuellement modalisés, relatifs à cette problématique : rencontres, partages, conversations, échanges, etc.

2. Voir plus loin, les enrichissements relationnels.

L'introduction de relations dans l'ouvrage est un acte délibéré et intentionnel des rédacteurs, conformément à nos positions de principe : Arcane ne dispose d'aucun algorithme d'analyse automatique de contenu, autre qu'un dispositif d'indexation lexicale optionnel.

Le choix et la définition des relateurs s'effectue, en principe, avant le début véritable du travail de rédaction, même s'il est possible à tout moment d'en ajouter ou de les modifier. C'est une phase essentielle de la méthodologie Arcane, comme la définition des types de sujets. C'est le moment, en effet, où le responsable de l'édition doit expliciter, conceptualiser et déclarer de manière opératoire et rigoureuse son projet éditorial. D'autant que, dans la perspective d'échanges dynamiques entre des ouvrages portant sur le même domaine et organisés en réseaux, ces modélisations doivent être compatibles et cohérentes.

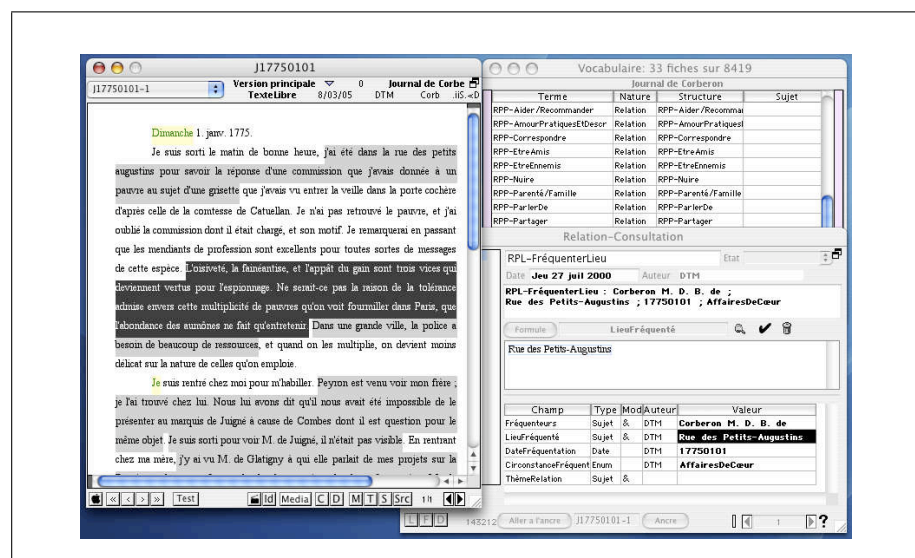


Figure 2. Une relation et son ancre.

Les arguments de type sujet, document ou relation constituent des arcs doublement orientés de la structure hypergraphique, et composent des liens dynamiques entre ces objets et la relation.

La figure 2 montre la fiche d'une relation, avec ses arguments typés et signés, son ancre dans un document, et quelques relateurs définis dans l'ouvrage.

Les utilisateurs font un grand usage des relations pour le repérage d'informations dans les documents ; il faut cependant noter la réticence qu'ils éprouvent à manipuler des relateurs ayant des arguments de type relation : leur inexpérience à la représentation opératoire de structures récursives en est certainement une des raisons.

2.2.3. *Sujets, relations et modélisation des connaissances*

Arcane a été développé bien avant que la problématique du « web sémantique » ne fasse l'objet de recherches très actives. Le modèle sujets-relateurs peut cependant être considéré comme une contribution à cette problématique puisqu'il vise lui aussi la formalisation opératoire de connaissances dans le but de les rendre partageables et transmissibles. De ce point de vue, Arcane peut apparaître comme un outil d'aide à la production d'ontologies, permettant à des spécialistes de produire des concepts (les sujets) et leurs relations dans différents domaines disciplinaires, même si les relateurs n'ont pas été conçus dans la perspective de composer des algorithmes générateurs complets et décidables.

Quoiqu'il en soit, ce modèle sujets-relateurs pourra être à l'avenir utilisé par l'ingénierie du web sémantique, lorsque celle-ci sera effectivement opérationnelle, ne serait-ce que par sa transcription dans le langage RDF (Ressource Description Framework), ou même OWL (Web Ontology Language).

2.3. *Le système Arcane*

Dans son principe, le système Arcane est une spécification pour l'implémentation d'un système informatique capable de mettre en œuvre le paradigme ; mais nous identifions dans la suite de cet article, le système avec notre propre implémentation prototype et opérationnelle. Le développement ouvert d'une version plus « industrielle » est aujourd'hui affaire d'opportunités et de ressources.

2.3.1. *Un système opérationnel de production et de recherche*

Le développement d'Arcane a été conditionné par l'impératif de proposer rapidement, avec des ressources limitées, un instrument de production efficace, adapté à des utilisateurs habitués aux pratiques éditoriales traditionnelles, et sans réelle culture informatique autre que le traitement de texte ; en même temps, il devait être suffisamment ouvert pour servir d'outil de recherche, et garantir que les éditions qu'il permettrait de réaliser seraient adaptables aux nouvelles technologies de la publication, telles qu'elles se dessinaient au début des années 1990.

2.3.2. *Un système modulaire*

L'architecture du système est issue d'une volonté *a priori* de concevoir un système modulaire et autonome, facile à mettre en œuvre, misant sur l'interopérabilité plutôt que sur la centralisation : une expérience antérieure de coopération avec des professionnels des sciences humaines et sociales [Lochard95a et b] nous avait convaincu que c'était la condition nécessaire pour que des utilisateurs engagent leurs propres projets éditoriaux ou de recherche, souvent confidentiels, avec un système encore prototype.

Le système Arcane est donc un ensemble de trois modules logiciels exécutables de manière autonome sous MacOS et Windows, articulés autour d'une même base de données qui contient la totalité de l'ouvrage édité : *un module d'écriture et d'enri-*

chissement pour les rédacteurs, en version mono-utilisateur ou client-serveur intranet, *un module d'administration* pour définir l'architecture des données, paramétrer l'environnement de travail, et produire les publications, et *un module web serveur* pour publier automatiquement des livres électroniques Arcane sous la forme de sites web dynamiques. Ces modules ont été développés à l'aide d'un générateur d'applications proposant sous forme de plugins des bibliothèques de gestion de base de données et de traitement de documents textuels, image et vidéo numérique.

Pour des raisons de pérennité, l'information de la base est exportable dans sa totalité sous forme de commandes SQL et dans un format XML.

2.3.3. *Un système d'édition original*

A notre connaissance et toute proportion gardée, Arcane n'a pas d'équivalent actuellement, si on prend en compte la mise en œuvre et la convivialité comme critère de comparaison, et si on ne tient pas compte des probables systèmes développés en interne dans des entreprises. Ce constat, qui nous étonne encore (et nous paraît de mauvaise augure), s'explique principalement par trois raisons.

La première tient au fait que la plupart des systèmes d'édition effectivement opérationnels sont des systèmes de publication³ et non des instruments d'écriture. C'est par exemple le cas de SPIP, LODEL ou Cyberdocs, spécialisés dans la « mise en ligne » d'articles et de documents issus des traitements de texte conventionnels : si la publication est électronique, les documents à publier ne le sont pas ; en outre, ces systèmes ne permettent pas de produire d'éditions papier.

En pratique, on peut distinguer un système de publication d'un système d'écriture électronique par la réponse donnée aux questions suivantes :

– « Le système nécessite-t-il, de fait, un travail préalable sur un traitement de documents (traitement de texte, grapheur, tableur) » : c'est oui pour un système de publication et non pour un système d'écriture qui est une alternative à ces traitements de documents ; on ne tient pas compte ici des traitements de texte intégré aux systèmes de publication en ligne qui ne permettent pas un travail rédactionnel effectif, compte tenu des faibles capacités actuelles des navigateurs en cette matière et/ou de la lenteur des échanges avec les serveurs web.

– « Qui sont les ouvriers principalement au commandes de ce système ? » : dans un système de publication, il s'agira plutôt d'ingénieurs des STIC, dans le cas d'un système d'écriture, des auteurs eux-mêmes (ou de leurs scripteurs) ; une forme alternative à cette question : « Quelle est la destination essentielle des ressources financières nécessaires au fonctionnement du système ? » : fonctionnement et salaire pour un système de publication, aide aux auteurs dans un système d'écriture.

3. Il ne faut pas confondre système de publication électronique et publication électronique, même si les composants informatiques sont souvent les mêmes (par exemple MySQL pour sélectionner les documents, et PHP pour les mettre au format HTML) : un système de publication est un outil électronique destiné à produire des publications.

– « Le système a-t-il recours à une base de connaissances pour l’indexation, (sachant qu’un mono fichier d’index n’en constitue pas une) ? » : c’est indispensable dans un système de production de documents électroniques.

– « Le système autorise-t-il un travail d’auteur sans accès à internet ? » : les systèmes de publication actuels ne le permettent pas ; et affirmer qu’aujourd’hui le travail « off-line » est inutile est un point de vue que notre expérience contredit.

La seconde raison tient au fait qu’un très grand nombre de systèmes qui implémentent des dispositifs d’écriture électronique ne peuvent pas être proposés aux utilisateurs visés par Arcane, soit parce qu’ils ont été développés pour des travaux de recherche « en interne », à l’occasion d’une thèse par exemple, soit parce qu’ils ne sont pas suffisamment aboutis. Les difficultés sont en effet nombreuses pour passer d’un prototype expérimental à un prototype opérationnel convivial adapté aux rédacteurs : il faut intégrer un système de gestion de base de données et des traitements de documents conviviaux, puis trouver des utilisateurs prêts à expérimenter en grandeur réelle des systèmes inhabituels. On peut citer comme exemple Porphyry (www.porphry.org), un prototype destiné au travail sur des corpus documentaires numérisés en archéologie, ou encore le projet HyperNietzsche (www.hypernietzsche.org), un système en développement prévu pour permettre « à une communauté savante délocalisée de travailler de façon coopérative et cumulative, et de publier les résultats de son travail sur Internet ». Un autre cas exemplaire est le poste de lecture initialement développé pour la Très Grande Bibliothèque (Grif SGML Editor, Thot) avec lesquels Arcane a de nombreux points communs, toute proportion gardée. Mais leur grande dépendance aux importantes ressources nécessaires à leur développement les rendaient fragiles : elle leur a été fatale, dès que les financements dont ces projets avaient bénéficié se sont portés ailleurs.

Enfin, une dernière raison tient à la généricité d’Arcane, conséquence du modèle de document implémenté et de la technologie SGML employée (il est né avant que XML ne s’impose comme standard dans l’industrie du logiciel) : les enrichissements, fondamentalement conçus comme des traitements logiques du document, sont enregistrés comme tels en dehors du document, et donc sans recours à un langage de balisage ; celui-ci peut être défini *a posteriori* par l’utilisateur pour produire directement toute sorte de documents algébriquement balisés, comme par exemple des documents XML, T_EX ou TEI. Arcane se comporte donc comme un outil de balisage générique *a posteriori* contrairement à d’autres systèmes de balisage conçus pour aider le rédacteur à partir d’un langage ou d’un format donné *a priori*.

2.3.4. Arcane et ses utilisateurs

L’installation et les mises à jour du logiciel n’ont jamais posé de difficultés aux utilisateurs habitués à la métaphore du bureau électronique : comme tout logiciel moderne, l’installation se fait par glisser/déposer. De plus, comme la totalité de l’ouvrage est enregistrée dans un seul fichier, il est assez simple de la transférer par internet pour des vérifications ou pour externaliser la production des publications.

La formation au logiciel lui-même est rapide puisque son utilisation est conforme aux habitudes prises avec les outils traditionnels de bureautique ; nous la pratiquons par demi-journées successives au fur et à mesure de l'avancée du travail d'édition : fonctionnement général, enrichissement, formulation de requêtes élaborées, éditions de travail, représentations graphiques.

La phase la plus délicate, mais la plus intéressante, est la définition et la prescription de l'architecture. Ce travail de modélisation formelle n'est pas familier aux chercheurs en SHS, et l'absence de toute culture mathématique rend quelquefois difficile la bonne gestion des relateurs d'ordre, l'écriture de scripts de représentation et l'expression logique dans les requêtes⁴. Pour les aider dans ces phases problématiques, nous proposons la collaboration d'un spécialiste de l'édition et du paradigme. Et pour tous les utilisateurs, les questions posées à propos d'Arcane deviennent très rapidement des questions relatives à l'édition, et non plus à l'instrument lui-même, si ce n'est pour solliciter le développement de procédures spécifiques.

Le lecteur intéressé par le point de vue des utilisateurs pourra consulter [Haffemayer02] ; les adresses électroniques de quelques uns d'entre eux sont également indiquées dans l'annexe.

3. Le document électronique Arcane

Dans le système Arcane, un document est un flot informationnel multimédia explicitement utilisé pour décrire, illustrer, commenter ou interpréter. C'est un objet interne du système, mais c'est également une unité logique d'information pour l'utilisateur, et il doit avoir la possibilité de l'identifier facilement en langue.

Comme nous l'avons déjà esquissé, la caractéristique fondamentale d'un document Arcane, qui en fait son originalité, est donnée par le principe de dualité :

document Arcane = support + ensemble d'enrichissements

Un document Arcane est ainsi un couple composé d'un « objet matériel inaltérable », le *support* du document, et d'un ensemble variable d'informations d'un type particulier, ses *enrichissements*.

Le support modélise l'aspect objectif du document, sa composante matérielle qui s'impose à tous les utilisateurs, lecteur et rédacteur ; les enrichissements sont les valeurs éventuellement subjectives que l'éditeur ajoute explicitement au document, pour lui donner une structure, construire ses représentations, composer un appareillage critique, un environnement paratextuel, et des interprétations.

4. C'est la raison pour laquelle nous militons activement pour l'introduction dans le cursus universitaire des sciences humaines et sociales, de formations de base sur les mathématiques discrètes, en particulier sur les structures d'ordre, les graphes, la logique et les langages formels.

Un document donné à lire ne s'identifie évidemment pas à son modèle Arcane, dès lors qu'il a une complexité sémantique suffisante : l'enrichissement et l'édition d'un document constituent *a priori* un processus sans fin ; sans oublier qu'il revient au lecteur de révéler le document par confrontation à son propre monde, processus qu'aucun rédacteur humain ou modèle formel ne peut contrôler.

L'expérience montre que les utilisateurs d'Arcane s'approprient facilement ce modèle ; d'abord parce qu'il est conforme aux habitudes prises avec le bureau électronique et les logiciels de traitement de texte conventionnels, mais aussi parce qu'il participe très naturellement à la rigueur du travail éditorial indispensable pour nourrir de manière cohérente d'importantes bases documentaires, sans pour autant contraindre la capacité créatrice des auteurs.

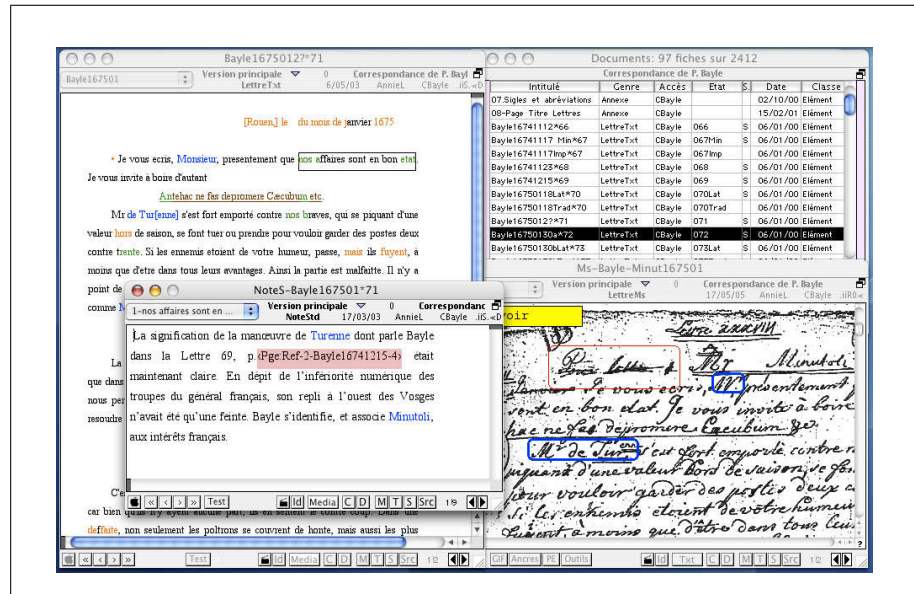


Figure 3. Documents Arcane : un texte, le manuscrit associé et un des codocuments.

3.1. Le support du document Arcane

Le support est l'aspect objectif du document, sa composante matérielle qui n'est pas affectée par la subjectivité de l'éditeur ni par celle des lecteurs ; dans le cas d'un texte, par exemple, le support sera identifié à la suite des caractères qui le composent.

Bien entendu, le support est la création d'un auteur et par là même peut être affecté d'un caractère subjectif ; cependant, pour des raisons pratiques, en particulier parce que les documents peuvent être des documents importés dans le système Arcane après avoir été fabriqués en dehors, l'aspect subjectif de la rédaction d'un support sera considéré dans Arcane comme encapsulé dans le support lui-même.

Une conséquence est que tout ce qui est peut être considéré comme relevant de la subjectivité du rédacteur fera l'objet d'un enrichissement ; l'inverse n'est pas toujours vrai : on utilisera aussi les enrichissements pour décrire la structure objective d'un document.

3.1.1. *Le support est inaltérable*

Un principe de la méthode Arcane énonce qu'un document devient enrichissable quand son support est considéré comme inaltérable ou pratiquement inaltérable⁵. La première raison est pratique : comme les enrichissements sont des objets extérieurs au support, il faut assurer la stabilité de l'ancrage et par conséquent du support ; la seconde raison est méthodologique et vise à étendre le principe d'encapsulation aux documents créés dans le système lui-même. On introduit ainsi explicitement deux moments caractéristiques dans l'histoire éditoriale d'un document : le temps de l'écriture de son support et le temps de son enrichissement ; ces deux états ne sont pas implémentés directement dans le système, mais le responsable de l'édition peut aisément les représenter par un champ descripteur associé au document, et s'en servir pour administrer l'édition.

3.1.2. *Le support est personnel ou distribué*

Un texte peut évidemment être créé avec Arcane, dans un module de traitement de texte intégré ; c'est, par exemple, le cas de toutes les annotations textuelles. Le support de ce document, original et signé, est enregistré dans l'ouvrage, de même que ses enrichissements.

La séparation effective du support d'un document et de ses enrichissements permet d'enrichir des documents dont le support est distribué par des banques de données institutionnelles accessibles par internet ou intranet, par exemple des documents iconographiques ou des manuscrits numérisés. Dans ce cas, le support est chargé dynamiquement pour visualiser le document, les enrichissements étant toujours enregistrés dans l'ouvrage. Une alternative plus commode et souhaitable à un tel *modus operandi* consiste à enregistrer dans l'ouvrage une copie du support, au besoin dans une version dégradée, d'effectuer le travail d'enrichissement sur cette copie et de ne télécharger l'original qu'en cas de nécessité, par exemple pour une publication.

3.1.3. *Le support est multimédia*

Le support d'un document est une suite linéaire d'unités informationnelles, quelle que soit la nature du média : un texte est considéré comme une suite de caractères « imprimables », une feuille de calcul comme une suite de cellules, une image comme

5. En pratique, il est possible de modifier le support d'un document, pourvu que cette modification soit marginale ou « structurelle » ; lorsqu'un document est à la fois rédigé et enrichi avec Arcane, les rédacteurs apprennent très vite à combiner efficacement l'écriture du support et son enrichissement.

une suite de pixels, un enregistrement sonore ou une vidéo Quicktime comme une suite temporelle.

La nature du média est propre au support ; les dispositifs pour effectuer les enrichissements, les enrichisseurs, prennent en compte le média seulement pour choisir la façon d'attacher l'enrichissement au support. Les procédures d'enrichissement sont ainsi rendues indépendantes du média du document, ce qui donne à l'outil Arcane une grande cohérence éditoriale.

D'une manière plus générale, les procédures de traitement des documents à la disposition des utilisateurs sont décrites indépendamment du média quand cela est fondé et possible. C'est le cas, par exemple, de la composition de documents virtuels, et de la recherche d'occurrences de sujets dans les documents.

3.1.4. *Le support est multilangue*

Le multilinguisme pris en charge dans le système Arcane est assez rudimentaire. Sa raison d'être est essentiellement de donner une solution commode sinon élégante à deux questions pratiques posées par les utilisateurs : rendre aisée la publication internet de l'ouvrage dans un site web international, et permettre l'édition critique de documents transcrits dans plusieurs langues.

La solution retenue est de considérer les traductions comme des versions différentes d'un même document, identifiées chacune par un support particulier, partageant le même ensemble d'enrichissements. Dans ce cas, le travail éditorial consiste à produire un autre support, à synchroniser les enrichissements sur ce support complémentaire, et éventuellement à produire une autre version linguistique pour les codocuments. Au moment de la consultation du document, le système charge le support *ad hoc* en fonction d'une préférence fixée par le rédacteur/lecteur.

Si cette solution est une réponse pragmatique à un problème de lecture, les problèmes des traductions peuvent être résolus selon la méthodologie Arcane habituelle : on distingue l'original de ses traductions par des documents différents, leurs liens étant traités par l'éditeur au moyen des dispositifs prévus pour gérer les liens entre documents.

3.2. *Sections d'un document*

Si le document est une unité logique d'information, son support est organisé en suite de sections.

La raison essentielle tient au confort de travail du rédacteur ; un document peut être long, et comme le travail éditorial avec Arcane n'est généralement pas linéaire, il est malaisé d'avoir à parcourir un document trop long. On découpe donc le document en une suite de sections-supports, suivant la structure naturelle du document.

La deuxième raison consiste, en effet, à s'adapter à la structure profonde de certains documents qui, se comportant comme des dossiers, sont explicitement composés d'unités logiques d'information (par exemple, un document composé des courriers échangés avec un correspondant, une lettre manuscrite ayant plusieurs feuillets). Arcane utilise lui-même cette architecture pour ses propres besoins ; par exemple, les codocuments attachés à un document par un enrichisseur codocumentaire donné, disons les notes explicatives, sont regroupés dans un seul (co)document associé au document source, chaque note constituant une section (les notes ne sont jamais très longues) : en ouvrant ce codocument et en le feuilletant section par section, on peut consulter (et enrichir) toutes les notes explicatives du document.

Une dernière raison est de faciliter l'édition critique à partir de manuscrits numérisés importés dans Arcane ; bien qu'il soit possible de réaliser une édition critique de documents images, il est en général nécessaire de faire une transcription du manuscrit pour le rendre lisible, et c'est à partir de cette transcription que l'appareillage critique est généralement composé. Le manuscrit contenant souvent plusieurs feuillets-images, on l'importe dans un document-image comprenant autant de sections qu'il y a de feuillets manuscrits ; puis on lui associe le document textuel de sa transcription composé lui aussi du même nombre de sections. Cette organisation permet de synchroniser, à la lecture, l'image originale et sa transcription ; elle permet également d'ancrer virtuellement certains enrichissements à la fois sur l'original et sur la transcription.

3.3. *Les enrichissements d'un document*

Les enrichissements sont des informations intentionnellement attachées par un rédacteur à un document, pour intervenir sur son contenu, sur ses multiples représentations et sur ses différences significatives proposées à la lecture.

Enrichir consiste à marquer logiquement tout ou partie du support, l'*ancree* de l'enrichissement, à l'aide d'un enrichisseur.

Plus précisément, d'une manière analogue au choix des attributs typographiques dans un éditeur de texte conventionnel, le rédacteur choisit l'ancree de l'enrichissement et lui applique un des enrichisseurs disponibles dans son environnement de travail ; il renseigne dans un formulaire les attributs de l'enrichissement s'il y en a, et attache un commentaire si c'est nécessaire. L'ancree est alors mise en forme en fonction des paramètres indiqués par l'enrichisseur associé.

L'ancree de l'enrichissement peut être le document complet, une section, ou encore, comme c'est généralement la cas, une séquence du support du document, suite contiguë d'unités du support : dans le cas d'une source textuelle, c'est la sélection habituelle de caractères, dans le cas d'une image, c'est un rectangle identifié à deux points, enfin dans le cas d'une bande sonore ou d'un film, c'est la séquence définie entre deux instants donnés.

Un enrichissement Arcane est par principe activable : une action de l'utilisateur sur son ancre provoque une réaction du système qui dépend du type de l'enrichissement et des valeurs données aux attributs. Si un commentaire multimédia est attaché à l'enrichissement, il est donné à lire, à voir ou à écouter au moment de l'activation.

Les enrichissements sont des objets aisément accessibles à l'utilisateur. Le rédacteur peut ainsi passer rapidement en revue toutes les ancres des enrichissements créées avec un enrichisseur particulier, retrouver des documents à partir d'une valeur donnée comme attribut d'un enrichissement, faire construire dynamiquement un document en organisant les ancres, ou les contextes des ancres, d'un ensemble d'enrichissements sélectionnés par requête.

3.3.1. Types d'enrichissement

Arcane distingue huit types d'enrichissements caractérisant chacun une fonction éditoriale.

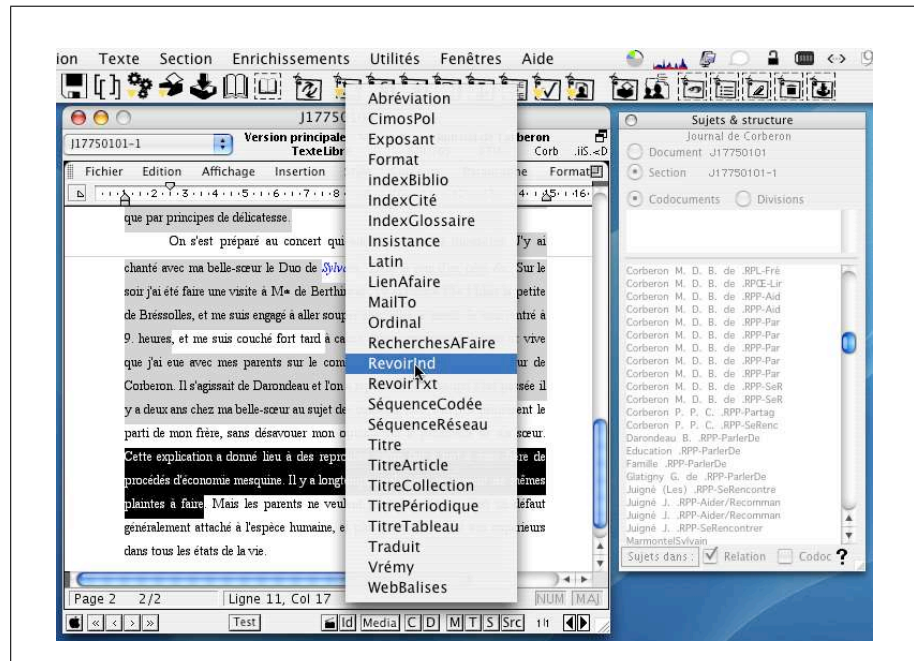


Figure 4. Enrichissement sémantique d'un document.

Etiquette sémantique (structure, représentation, signification) Le rédacteur attache à l'ancre une étiquette caractérisée par l'enrichisseur ; des attributs permettent d'affiner cet étiquetage. Son utilisation est multiple : marquer logiquement une séquence (Titre, Langue, URL, ...), attacher un état éditorial (Revoir, Césure possible, ...).

Paragraphe (structure, représentation, signification) Identique au précédent type, il est ancré obligatoirement sur un ou plusieurs paragraphes ; le marquage logique des paragraphes (Remarque, Exercice, Liste, Tableau, ...) est une procédure fréquente pour définir la structure d'un document.

Insertion multimédia (contenu) Cet enrichissement sert à composer le contenu d'un document à partir de tout ou partie d'un autre document de média compatible ; l'insertion est effectivement réalisée sur demande de l'utilisateur, ou automatiquement lorsque c'est nécessaire, pour produire une publication ou exporter le document, par exemple. On notera que l'élément à insérer fait partie du contenu du document, mais pas de son support.

Division (structure) On l'utilise pour structurer le document ou une liste de documents en une suite de subdivisions : chapitres, sections, sous-sections etc, par exemple pour produire la table des matières d'une publication. L'ancre est le titre de la subdivision. Chaque enrichisseur de division définit le style de l'ancre (le titre de la subdivision) et le style du corps de la subdivision.

Lien explicite (signification) Le rédacteur lie dynamiquement, éventuellement de manière réciproque, l'ancre à une cible qui peut être un document, une séquence, un enrichissement, la vue d'un sujet, ou encore une information extérieure à l'ouvrage. C'est l'enrichissement idéal pour supporter un commentaire et pour construire des parcours intentionnels de lecture dans une publication électronique.

Le qualificatif « explicite » tient au fait que d'autres types d'enrichissements se comportent comme des liens : leur activation déclenche la présentation d'un ou plusieurs objets cibles.

Codocument (signification) En pratique, cet enrichissement associe à une ancre une nouvelle section d'un document spécialisé, propriété du document source ; c'est la note de bas de page.

Le nombre d'enrichisseurs codocumentaires que le responsable de l'édition peut mettre à la disposition du rédacteur n'est pas limité : note de commentaire, note critique, note de travail, d'administration, indication et solution d'exercice, etc ; compte tenu de la variété de ces enrichissements, nous préférons la dénomination « codocumentation » à « annotation ».

Cet enrichissement, comme la plupart des autres, s'applique aux documents indépendamment du média : on peut ainsi codocumenter une image, une bande sonore, ou encore un film. D'un autre côté, les codocuments sont eux aussi multimédia : parmi les nouvelles pratiques de l'édition, on pourrait, par exemple, recommander que les notes d'un certain type soient écoutées plutôt que lues.

L'enrichissement codocumentaire a des points communs avec le lien explicite : ce sont tous les deux des liens d'une ancre vers la vue d'un objet. La différence essentielle, outre celle de la mise en œuvre, tient au rapport de dépendance existentielle entre la source et la cible : quand on supprime un document, les objets cibles des liens ancrés sur ce document ne sont pas affectés, alors que, du fait de l'exclusivité de la relation entre un document et ses codocuments, tous

les codocuments du document supprimé sont automatiquement effacés par le système.

Dans Arcane, la notion de codocumentation ne s'applique pas seulement aux documents, mais également aux sujets et aux relations ; ainsi les documents dont l'existence est liée à celle d'un sujet ou d'une relation (bibliographie d'une personne, commentaire d'une relation, par exemple) sont de classe codocument : ils sont supprimés avec l'objet qui justifie leur existence.

Relation (signification) Un enrichissement de ce type consiste à associer à l'ancrage une relation Arcane ; il généralise la procédure d'indexation, et permet la formalisation du contenu ; nous y reviendrons.

Accès documentaire (signification) Sa fonction est essentiellement de définir une séquence d'un document multimédia comme un élément documentaire directement accessible ; cet élément peut concerner un sujet (élément biographique, commentaire d'une personne sur une autre personne dans un document, image d'un sujet dans une photo de groupe ...) auquel cas il est indexé au sujet pour être accessible à partir de la fiche de celui-ci, au même titre qu'un codocument de sujet ; il peut également être intégré à une liste de points d'entrée général dans l'ouvrage, facilement identifiable par le lecteur, comme le sont par exemple les « Frequently Asked Questions » dans une documentation, ou les « Quiz » dans un ouvrage pédagogique. Ces éléments documentaires sont des candidats naturels pour la production des documents virtuels

3.3.2. *Enrichisseur*

Les enrichisseurs sont les mécanismes de traitement des enrichissements ; ils sont créés par instanciation d'un des huit types précédents implémentés dans la méta-architecture Arcane. Lorsqu'un enrichisseur est prescrit à un couple (genre de document ; groupe d'utilisateurs) il est mis à la disposition de tous les rédacteurs du groupe dans leur environnement de travail, pour traiter les documents de ce genre (cette notion est présentée plus loin).

Un enrichisseur est librement créé, décrit et prescrit par le responsable de l'édition. Il fixe les options correspondantes au type d'enrichissement, les styles graphiques, typographiques et les formats de mise en forme à appliquer sur les ancres dans le document de travail, il définit les attributs de l'enrichissement que le rédacteur devra renseigner, et il spécifie le traitement à appliquer à l'enrichissement au moment de la production d'une publication ou de l'exportation dans un format donné, le balisage notamment.

Ce mode de mise en œuvre des enrichissements permet à l'éditeur de composer un environnement de travail cohérent, efficace et adapté à chaque rédacteur ou groupe de rédacteurs, en particulier par l'usage de chartes graphiques, qu'il peut, bien entendu, modifier à tout moment.

4. Le document dans son environnement Arcane

L'écriture électronique inscrit le document dans un réseau plus ou moins dense d'interconnexions avec les autres objets de l'ouvrage. La *classe* d'un document précise sa position relative dans cette architecture hyper-graphique. Son *genre* sert à regrouper logiquement les documents d'une même classe, pour faciliter leur repérage par l'utilisateur, mais aussi pour leur attribuer certaines propriétés éditoriales communes, dans le but de favoriser la cohérence interne de l'ouvrage : champs descripteurs, ensemble d'enrichisseurs disponibles, formats de représentation, etc.

4.1. Documents matériels et virtuels

Les documents multimédias sur lesquels les rédacteurs effectuent des enrichissements sont les *documents matériels* : Arcane enregistre directement ou indirectement leur support dans la base, ainsi que les enrichissements. Les *documents virtuels*, *dynamiques* dans la terminologie Arcane, sont produits par application d'un algorithme *ad hoc* interne au système ou programmé, à des informations de l'ouvrage ; ces documents électroniques ne sont pas enregistrés (sauf si le rédacteur en fait explicitement des documents matériels) et ne peuvent donc pas être enrichis. Les documents matériels enrichis par des insertions constituent une forme hybride puisque leur contenu est également le produit d'un calcul : ce sont les *documents mixtes*.

4.1.1. Les documents matériels.

Les documents matériels sont les documents enrichissables de l'ouvrage. Arcane distingue quatre classes : les documents autonomes, les codocuments de sujet, de document ou de relation.

Document autonome Ce sont les documents introduits dans l'ouvrage pour leur intérêt propre, qui justifient leur existence par eux-mêmes. Cet article est un exemple de document autonome ; les manuscrits et les transcriptions cités plus haut en sont d'autres.

Codocument C'est un document multimédia dont l'existence dépend d'un autre objet Arcane, sa *source* : la suppression d'un objet source entraîne automatiquement la suppression « en profondeur » de tous les codocuments qui lui sont directement et indirectement attachés. Cette propriété est souvent le critère qui permet à l'éditeur, en cas d'hésitation, de choisir la classe à donner à un document à créer. D'un point de vue opératoire, un codocument est toujours créé par l'intermédiaire d'un enrichissement codocumentaire.

Codocument de document C'est, nous l'avons vu, une généralisation fonctionnelle de l'annotation. Rappelons que l'architecture Arcane ne limite pas le nombre de genres de codocuments appelés à codocumenter un genre de document. Comme par ailleurs un codocument est un document Arcane, il peut être lui-même codocumenté ; la profondeur de la codocumentation n'a pas d'autre limite que celle de pouvoir clairement spécifier en langue les différents niveaux. Un travail est

actuellement en cours pour classer les codocuments en fonction de la manière dont ils doivent être présentés au lecteur après activation de leurs ancrés ; cette présentation dépend du média utilisé et de la fonction codocumentaire dans le processus de lecture : on ne présente pas de la même manière une remarque succincte « post-it », une explication complémentaire détaillée, une indication pour un exercice ou sa correction.

Codocument de sujet Il est utilisé pour décrire, commenter ou illustrer un sujet : biographie et portraits d'une personne, bibliographie d'un ouvrage, note lexicale, définition d'une notion, etc ; la suppression du sujet entraîne là aussi la suppression de tous ses codocuments. Les codocuments d'un sujet sont non seulement accessibles à la lecture par la fiche du sujet, mais le lecteur peut également les repérer comme n'importe quel autre document.

Codocument de relation Il sert à commenter la relation. Cette codocumentation est effectivement utilisée lorsque la relation est une information centrale de l'ouvrage. C'est surtout le cas des relations autonomes, c'est-à-dire non ancrées sur un document ; pour les autres, c'est l'enrichissement relationnel qui doit être commenté plutôt que le relation.

La classe d'un document est implicitement utilisée par le lecteur pour consulter sans difficulté un document : si c'est un document autonome, le lecteur a de multiples possibilités pour le sélectionner : par son genre et/ou son intitulé et/ou son contenu, etc ; si c'est un codocument, il va le consulter à partir de l'objet source : le sujet, le document ou la relation.

4.1.2. *Les documents virtuels*

Comme nous l'avons dit, le modèle Arcane de l'écriture électronique n'est pas tant la production de documents virtuels que celle de documents électroniques, matériels ou virtuels ; cela n'empêche pas ces derniers d'avoir une place importante dans le système Arcane.

La richesse combinatoire des documents virtuels ne tient pas aux algorithmes de composition, il y en a relativement peu, mais à la grande quantité et à la variété des éléments d'information qui peuvent être utilisés (des milliers de sujets, de relations et de documents, des dizaines de milliers d'enrichissements). D'autre part, l'ensemble des éléments à composer est une sélection spécifiée directement ou indirectement par l'utilisateur dans le langage Arcane, et non des objets « automatiquement » sélectionnés dans une approche déclarative, à l'aide de propriétés et de règles mises en œuvre à partir d'une déclaration d'intention de l'utilisateur (voir par exemple [Iksal02]) ; ces deux approches pourraient d'ailleurs s'avérer complémentaires, si on voulait munir Arcane d'un module sophistiqué d'aide à la formulation de requêtes. Ainsi, dans l'état actuel du modèle Arcane, la production de documents virtuels est directement liée au choix de l'architecture, en particulier des enrichisseurs, dont la pertinence et l'efficacité est affaire de pratique.

Pour l'instant, les documents virtuels sont essentiellement utilisés comme documents de travail. Le principal document est l'assemblage d'une sélection d'enrichissements, qui consiste en une « concaténation » des ancres ou du contexte des ancres des enrichissements. Le résultat est un texte électronique lié dynamiquement aux sources de l'assemblage, un diaporama d'images, ou même un document QuickTime composé à la volée ; l'utilisateur peut ainsi « lire » sur un même document toutes les séquences indexées par un sujet, par un enrichisseur donné, ou étiquetté à des fins éditoriales.

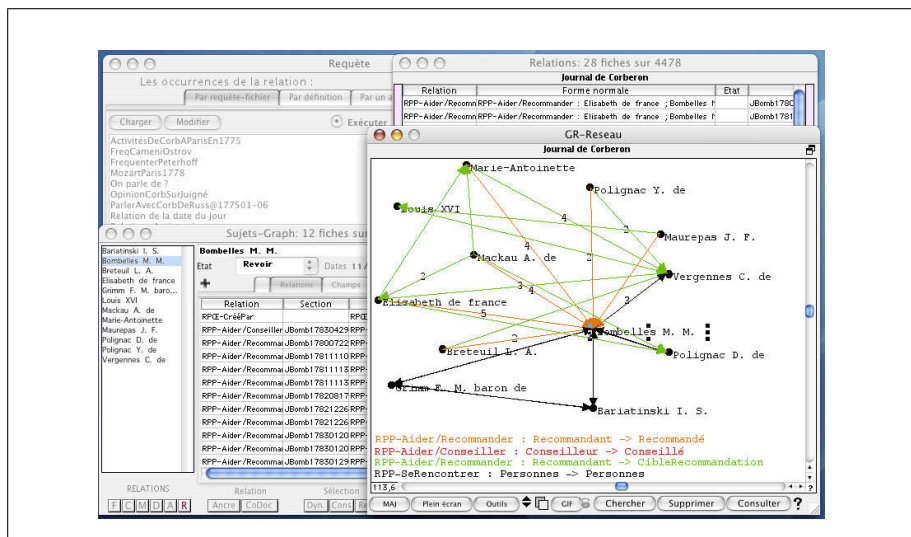


Figure 5. Représentation graphique de relations de sociabilité.

Les représentations graphiques (ou cartographiques) de sujets ou de relations sont également des documents virtuels très prisés ; la figure 5 montre un document virtuel image, composé à partir d'une sélection de relations concernant un sujet ; là encore, les utilisateurs sont handicapés dans l'écriture des scripts de représentation par leur incompétence en matière de graphes, ce qui rend cette écriture souvent cahotique et nous amène à les rédiger pour eux. La difficulté est d'ailleurs du même ordre pour les représentations statistiques.

On peut également indiquer comme documents virtuels, les index cumulatifs automatiques, les tables des matières, ainsi que toutes les représentations paramétrables de documents sous divers formats (DOC, RTE, HTML, JPEG, etc.).

4.2. Le document comme sujet

Si un document Arcane est naturellement un flot informationnel, un vecteur de la lecture, il peut également exister en tant que sujet Arcane, si le document est un sujet d'intérêt de l'ouvrage ; c'est par exemple le cas d'une édition de correspondances,

dans laquelle une lettre est à la fois un sujet à décrire, une image-manuscrit et un texte-transcription.

Comme la relation de codocumentation n'est pas satisfaisante (le sujet et le document constituent une même entité, différemment représentée, et ne sont pas dans une relation de dépendance), Arcane propose un mécanisme d'encapsulation-abstraction qui donne au lecteur des outils pratiques pour manipuler l'entité par toutes ses instances Arcane à la fois ; ainsi par exemple, les champs descripteurs du document-sujet peuvent être réunis aux champs descripteurs du document-document.

4.3. *Indexation et représentation du contenu*

Bien que les documents Arcane peuvent être considérés comme des documents conventionnels au même titre que les documents produits par les logiciels standard du marché, le paradigme les inscrit dans un monde de connaissances explicites et intentionnelles, formellement présentes dans l'ouvrage : l'environnement sémantique fait donc partie du document et doit être publié avec lui : si cet environnement est créé ou recréé *a posteriori*, après sa publication, on a affaire, de notre point de vue, non pas à la publication d'un document mais à celle d'un support. D'autre part, comme le modèle de publication implicitement spécifié dans le web conduit nécessairement à une explosion quantitative de l'offre, il est essentiel pour un éditeur qui souhaite rencontrer des lecteurs dans la perspective d'un contrat de lecture, de faciliter le repérage de ses publications, en particulier par des marqueurs sémantiques pertinents. Et il nous paraît beaucoup plus judicieux qu'il s'en occupe lui-même, plutôt que d'abandonner ce travail à des automates, aussi puissants soient-ils. Enfin, une de nos ambitions étant de faire communiquer dynamiquement les livres électroniques entre eux, plus la richesse combinatoire des mondes de connaissances enregistrés dans ces livres sera grande, plus utile et intéressante sera cette communication.

L'enrichissement sémantique est donc une opération fondamentale de l'écriture avec Arcane, même si certains utilisateurs la limitent encore à la production d'index cumulatifs automatiques dans leurs projets d'édition papier.

Dans Arcane, faire un enrichissement sémantique consiste à attacher à l'ancre un sujet de l'ouvrage, indexation simple, ou une relation, indexation relationnelle.

4.3.1. *Indexation simple*

Ce mode d'indexation revient à énoncer que l'ancre de l'enrichissement « parle du sujet indexé » ; on peut la voir comme la forme la plus triviale de l'enrichissement relationnel.

Le système crée un lien interne réciproque entre le document et le sujet : depuis l'ancre, le lecteur peut activer la présentation du sujet, et inversement, l'ancre peut être directement consultée dans une vue spécialisée du sujet, rendant le document directement accessible.

Tout enrichissement est potentiellement une indexation : il suffit que le paramétrage de l'enrichisseur prévoie un attribut de type Sujet, et que le rédacteur renseigne cet attribut. Un exemple classique est l'étiquette « Titre ... » qui permet à la fois d'extraire de l'ouvrage toutes les occurrences de cet enrichisseur, de prescrire le style italique pour l'ancre si elle est textuelle, et de lui associer un ouvrage enregistré comme sujet.

4.3.2. *Indexation relationnelle*

Pour ancrer sur le support d'un document une occurrence d'un relateur, le rédacteur choisit l'enrichisseur relationnel associé, et renseigne dans un dialogue les arguments définis par ce relateur, ainsi que les éventuels attributs de l'enrichissement ; l'ancre est mise en forme selon le style typographique défini dans l'enrichisseur, ce qui permet de distinguer visuellement les différentes relations associées à un document.

D'un point de vue opératoire, une relation établit un ensemble de liens dynamiques, implémentés ou calculables, entre ses objets-arguments ; d'une vue d'un sujet par exemple, le lecteur peut consulter la liste des relations dans lesquelles le sujet est un argument ; il peut également programmer des représentations graphiques dynamiques d'un ensemble de relations, les utiliser pour exécuter des requêtes très précises ou introduire du vocabulaire dans l'ouvrage afin d'accroître la combinatoire sémantique de l'ouvrage et faciliter sa lecture ; il peut aussi les exporter dans un format standard pour des traitements complémentaires, logiques par exemple.

C'est le geste intentionnel du rédacteur qui augmente la combinatoire sémantique du document et donc de l'ouvrage, et non le mécanisme d'enrichissement relationnel qui n'est qu'un mécanisme formel ; il est donc essentiel que le mode d'emploi de chaque relateur soit très clairement établi afin d'assurer la cohérence sémantique de l'ouvrage, et de garantir le contrat éventuel de lecture, surtout si l'édition est collective.

4.3.3. *Environnement sémantique explicite d'un document*

Les indexations simples et relationnelles ancrées sur un document définissent un ensemble de sujets et de relateurs qui constituent en quelque sorte l'environnement sémantique explicite du document. Sur la figure 4 on peut ainsi voir la liste activable des occurrences des sujets indexés dans le document, avec les relateurs éventuels.

4.3.4. *Relateurs universels*

Une part importante de notre travail expérimental avec les équipes d'éditeurs a porté sur les relateurs. La question était de comparer les relateurs utilisés dans des projets éditoriaux « compatibles » (l'étude de la sociabilité à partir de correspondances, par exemple), et d'établir des relateurs communs (en forme et en mode d'emploi).

L'objectif est de les normaliser pour permettre non seulement la communication dynamique entre les ouvrages électroniques Arcane, mais également de les utiliser

dans le futur format du web sémantique, pour instancier les méta-données associées à un document HTML.

4.3.5. *Relation autonome*

Une relation peut être créée en dehors de tout document afin d'établir des hiérarchies entre les sujets ou de les organiser (sortes, sous-sujet, renvois), ou d'exprimer une relation objective entre des sujets (appartenance, relation biographique et plus généralement chronologique...).

5. Publication des documents

Dans le système Arcane, la publication est conduite par un script qui permet à l'utilisateur de déterminer :

- l'ensemble des documents ou des sujets à publier,
- les enrichisseurs (ou les enrichissements) à prendre en compte,
- et le paramétrage global de la publication ainsi que l'agencement des différentes informations.

5.1. *Publication papier*

Les impressions de qualité sont confiées à \TeX ; la publication revient dès lors à produire un document balisé \TeX , qui est directement compilable dès que les rédacteurs travaillent avec suffisamment de précision.

Les index cumulatifs sont composés à partir des indexations, simples ou relationnelles : un champ *ad hoc*, généralement la forme normalisée enrichie, est utilisé pour produire les vedettes ; les relations autonomes sont utilisées pour hiérarchiser les index.

5.2. *Publication électronique*

Le système Arcane étant composé d'une base de données sur laquelle opèrent les différents modules logiciels Arcane, une première option consiste à publier et à distribuer la base de données réduite aux informations publiables, équipée d'un module de lecture (le module de rédaction bridé).

La seconde option consiste à diffuser dynamiquement la publication sous la forme d'un site web distribué par le module web serveur. Cette solution est particulièrement bien adaptée à la publication progressive, puisqu'elle permet à l'éditeur de travailler en même temps sur l'ouvrage, soit par intranet avec le module éditeur, soit par internet au moyen de formulaires HTML spécialement conçus à cet effet ; dans ce dernier cas, le

travail d'édition se limite à des opérations élémentaires, compte tenu des performances actuelles des navigateurs.

La publication web de l'ouvrage est considérée comme la production de documents virtuels partiellement conduite par le lecteur, et, si le travail d'édition a été convenablement mené, elle ne requiert aucun travail supplémentaire. En effet, l'architecture Arcane définit non seulement les objets du monde, mais également des vues de représentation de ces objets ; à partir d'un modèle de page le serveur web traduit des vues en blocs d'informations HTML identifiés, en laissant à la feuille de style préférée du lecteur le soin de faire la mise en page sur son navigateur ; un langage de composition-extraction permet en outre de composer des documents annexes à l'ouvrage. La figure 6 donne un aperçu du document de la figure 3, sans doute provisoire, publié dans la vue de consultation des documents.

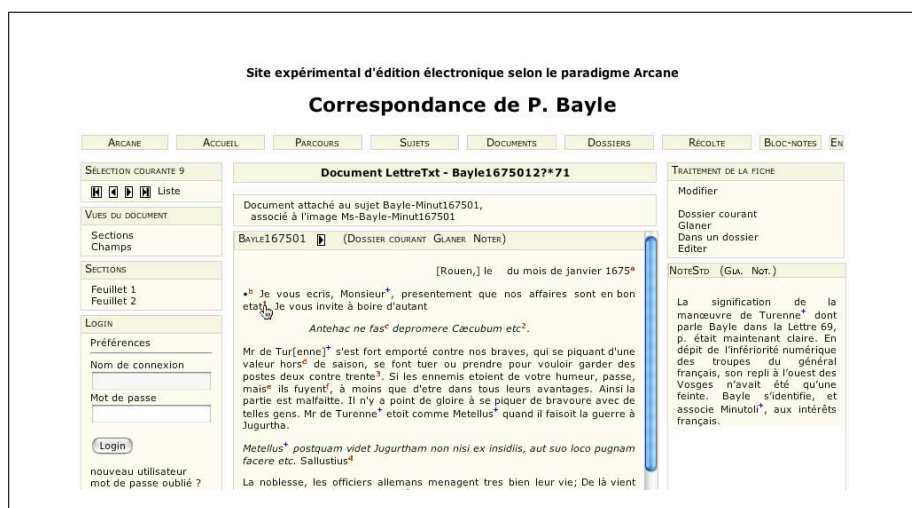


Figure 6. La lettre de la figure 3 publiée dans un livre web.

5.3. Publication balisée

Nous avons déjà dit que le système Arcane peut être utilisé comme un système de balisage algébrique de documents. Il « suffit » de créer dans l'architecture de l'ouvrage des enrichisseurs conformément à la structure et/ou la sémantique du langage de balisage, puis, avant l'exportation des documents dans ce format, de fixer la balise *ad-hoc*, éventuellement attribuée, de chacun des enrichisseurs. Cela suppose que la structure du langage de balisage soit compatible avec la méta-architecture Arcane qui a été conçue pour la production de documents électroniques originaux, plutôt que pour des opérations de « rétro conversion » de documents déjà publiés, par exemple pour décrire dans une transcription la forme du manuscrit d'origine. La publication balisée permet également de fournir des données à des serveurs de publication utilisant des

systèmes de gestion de bases de données standard, au format SQL par exemple. Ces données sont alors exportées dans des tables définies selon le modèle de la méta-architecture Arcane : sujets, champs des sujets, relations, arguments des relations, sections, enrichissements des sections etc. Cette solution a été prévue dès l'origine pour assurer la pérennité du contenu des ouvrages édités.

6. Perspectives

Le modèle Arcane visait dès l'origine des perspectives que nous n'avons pas évoquées, ou alors très brièvement. Certaines d'entre elles nécessitaient la réalisation préalable d'ouvrages comme matériau de travail, les autres sont encore aujourd'hui à l'état de projet. Nous indiquons ici celles qui ont un rapport direct avec l'écriture et la lecture de documents électroniques.

6.0.1. *Les parcours de lecture*

Les parcours sont des suites de vues d'objets organisées selon un scénario intentionnel. Les premiers essais expérimentaux que nous avons réalisés démontrent le grand intérêt de ce dispositif multimédia, en particulier comme forme d'exercice pratique de lecture dans un environnement pédagogique. L'essentiel du travail consistera à concevoir un interface efficace et convivial.

6.0.2. *L'échange dans un réseau d'ouvrages*

A l'origine, le développement d'Arcane s'inscrivait dans un projet de recherche visant à réaliser un atlas européen de correspondances de l'époque moderne, constitué d'un ensemble d'ouvrages électroniques capables de communiquer entre eux. L'objectif est de donner la possibilité à un lecteur /rédacteur d'accéder dynamiquement à une information documentaire dans un autre ouvrage, susceptible de la lui fournir. Le travail des éditeurs nous permet de disposer aujourd'hui de volumineuses bases de données portant sur des domaines voisins, éditées dans des architectures similaires ; la mise en œuvre de ce travail va surtout dépendre des ressources que nous pourrions lui affecter.

6.0.3. *La lecture « en langue »*

La création d'un sujet ou d'un document dans l'ouvrage est automatiquement enregistrée dans un dictionnaire, avec son intitulé comme vedette. Ce dictionnaire contient initialement tous les termes du méta-langage Arcane, et les éléments des structures prescrites dans l'architecture (types de sujets, champs descripteurs, relateurs, arguments, etc). Les rédacteurs peuvent y ajouter des formes lexicales équivalentes au nom des sujets, et des définitions créées à partir des relations ou des champs descripteurs par des techniques d'unification. La fonction de ce dictionnaire est évidemment de faciliter les recherches dans la base de données. Mais il a également été conçu pour permettre d'explorer l'ouvrage « en langue », par l'écrit ou à la voix, à partir d'un langage formel convivial qui reste à décrire.

7. Bibliographie

- [Berners01] Berners-Lee, T., Hendler, J., Lassila, O. « The Semantic Web : A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities », *Scientific American*, May 2001.
- [Bush45] BUSH, Vannevar. « As We May Think » *The Atlantic* 176.1, July 1945, p. 101-108.
- [Chartier95] Chartier, Roger, *Forms and meanings : Texts, performances and audiences from Codex to Computer*, Philadelphia, University of Pennsylvania Press, 1995.
- [DVP] Documents Virtuels Personnalisables 2002, <http://iasc.enst-bretagne.fr/DVP2002/>
- [FalquetAl04] Gilles Falquet, Luka Nerima, Jean-Claude Ziswiler, *Utilisation des ontologies dans les environnements d'écriture collaborative d'hyperlivres, expériences et perspectives*, Jacqueline Bourdeau, Michel Crampes (Éds.) *Que signifie l'ontologie pour les EIAH ? Revue STICEF*, Volume 11, 2004.
- [Haffemayer02] Haffemayer S., *Arcane, un paradigme pour l'analyse textuelle*, <http://www.uottawa.ca/academic/arts/astrolabe/articles/art0018.htm>, Arcane.htm
- [Iksal02] Sébastien Iksal, *Spécification déclarative et composition sémantique pour des documents virtuels personnalisables*, Thèse ENST Bretagne, 2002
- [Landow92] Landow, George P., *Hypertext. The Convergence of contemporary critical Theory and technology*, Johns Hopkins University Press, 1992.
- [Landow94] Landow, George P., *Hyper/Text/Theory*, Johns Hopkins University Press, 1994.
- [Landow97] Landow, George P., *Hypertext 2.0 : The Convergence of Contemporary Critical Theory and Technology*, Baltimore, Johns Hopkins University Press, 1997.
- [Lochard95a] Lochard, Eric Olivier, « A propos de TOPOSATOR », colloque SATOR de Winnipeg, *Utopie et fictions narratives*, éd. Michel Bateau et Santé Viselli, Parabasis, Altapress, 1995, p. 257-268.
- [Lochard95b] Lochard, Eric Olivier, « TOPOSATOR et ARCANE : la nouvelle technologie dans le contexte des thesaurus et des éditions critiques », *Literary and linguistic computing*, Oxford University Press, vol. 10, n°2, 1995, p. 151-164.
- [LochardTaurisson01] Lochard, Eric Olivier, Taurisson, Dominique, « Le monde selon Arcane », *Le Document au XXIe siècle, Cahiers Gutenberg*, n° 39-40, mai 2001, p. 89-105.
- [LochardTaurisson02a] Lochard, Eric Olivier, Taurisson, Dominique, « " The World According to Arcane ", An operating Instrumental Paradigm for scholarly Editions », *Perspectives of Scholarly Editing/Perspektiven der Textedition*, ed. by Bodo Plachta et H. T. M. van Vliet, Berlin, Weidler Buchverlag, 2002, p. 151-162.

- [**LochardTaurisson02b**] Lochard, Eric Olivier, Taurisson, Dominique, « Correspondances, réseaux, édition électronique », Actes du colloque *La Plume et la toile. Pouvoirs et réseaux de correspondance dans l'Europe des Lumières*, Arras, 26-27 octobre 2000, éd. P.-Y. Beaurepaire, avec une introd. de Daniel Roche, Artois Presses Université, 2002, p. 171-192.
- [**Nelson65**] Nelson, Theodor H. *The Hypertext, Congress of the International Federation for Documentation Abstracts*, October 10-15, 1965, Washington, DC. 80.
- [**Pédauque03**] *Le document comme contenant, contenu et médium. Les reformulations du numérique*, Roger T. Pédauque; http://archivesic.ccsd.cnrs.fr/sic/_00001115.html
- [**Pédauque04**] *Le texte en jeu. Permanence et transformations du document*, Roger T. Pédauque; http://archivesic.ccsd.cnrs.fr/sic/_00001401.html
- [**Ranwez**] Sylvie Ranwez, *Composition automatique de documents hypermédia adaptatifs à partir d'ontologies et de requêtes intentionnelles de l'utilisateur*, Thèse Montpellier, 2000.
- [**RDF**] Resource Description Framework, <http://www.w3.org/TR/rdf-primer/>
- [**TEI**] Text Encoding Initiative, <http://www.tei-c.org/>
- [**Trudeau00**] Trudeau, Danielle, « L'Apport des nouvelles technologies au domaine de l'édition critique », *Les Etudes françaises valorisées par les nouvelles technologies de l'information et de la communication*, colloque international, Université de Toronto, 12-13 mai 2000.
- [**Vandendorpe99**] Vandendorpe, Christian, *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*, Montréal, Boréal et Paris, La Découverte, 1999.
- [**Vignaux00**] Vignaux, Georges, « L'Hypothèse du livre électronique », *Cahiers de médiologie*, n°10.

8. Annexe : Edition et travaux de recherche effectués avec Arcane

Arcane est aujourd'hui utilisé dans plusieurs programmes d'édition critique de correspondances du XVIII^e siècle, en particulier :

- Edition de la *Correspondance complète de Pierre Bayle* (dir. Antony McKenna, mckenna@univ-st-etienne.fr, Institut Claude Longeon UMR 5037-CNRS ; Fondation Voltaire, Oxford ; le volume III est paru en 2004, le volume IV est prévu pour 2005) ;

- *Inventaire de la Correspondance de Samuel Formey* (dir. Jens Häsel, haeseler@rz.uni-potsdam.de, Institut des Lumières, Potsdam ; Champion, Paris, 2003) ;

- Edition électronique du *Journal du Chevalier de Corberon* (dir. Pierre-Yves Beaurepaire et Dominique Taurisson, dominique.taurisson@ehess.univ-mrs.fr ; <http://egodoc.revues.org/corberon/>)

- Inventaire critique de la *Correspondance* du théologien genevois Jean-Alphonse Turretini (dir. Maria-Cristina Pitassi, maria-cristina.pitassi@ihr.unige.ch, Institut de la Réformation de Genève ; 4 volumes à paraître chez Champion, Paris, 2005)

D'autres chercheurs ou équipes d'auteurs l'utilisent pour leurs propres travaux de recherche ; on peut citer :

- *Correspondance de la Mission catholique d'Ecosse 1730-1829* (Dir. Clotilde Prunier, clotilde.prunier@univ-montp3.fr, Université de Montpellier 3 ; 25000 lettres manuscrites pour étudier le réseau de la minorité catholique écossaise en Europe) *Réseaux de sociabilité et correspondances : le rôle des femmes dans la diffusion du savoir et de la culture entre la France et la Nouvelle-France* (Dir. Julie Roy, julie.roy4@sympatico.ca, ALAQ/Université du Québec à Montréal)

- Projet du C.E.R.C.O.R. (Centre européen de Recherches sur les Congrégations et Ordres Religieux, Université de Saint-Etienne-CNRS ; Dir. Hélène Morin-Sauvade, helene.morin-sauvade@laposte.net) : Création d'une base de données scientifique et d'un site Web pour répertorier les anciennes fondations cisterciennes masculines et féminines en Europe (sources, bibliographie, cartes, tableaux de filiation, notices détaillées, etc.).

Trois étudiants en thèse (Université de Bordeaux 3 et Université de Nice) l'ont adopté pour leurs travaux de recherche et pour la rédaction de leurs thèses.

Enfin, à la rentrée, Arcane sera utilisé à titre expérimental comme outil de travail dans un cursus universitaire en licence d'histoire.